ELSEVIER Journal of Econometrics 64 (1994) 355–373

JOURNAL OF Econometrics

Coherency and estimation in simultaneous models with censored or qualitative dependent variables

Richard Blundell*,[a], Richard J. Smith[b]

[a] *Department of Economics, University College, London WC1E 6BT, UK*
[b] *Gonville and Caius College, Cambridge CB2 1TA, UK*

(Received March 1991; final version received May 1993)

Abstract

This paper concerns estimation and inference in simultaneous limited dependent variable or simultaneous qualitative dependent variable models. We discuss the implications of the coherency condition which is required to ensure a unique implicit reduced form in such nonlinear models. A number of model specifications are examined, including the *Simultaneous Probit*, *Simultaneous Tobit*, and *Simultaneous Generalised Selectivity* models, and we consider the usefulness of a conditional maximum likelihood (CML) methodology. We obtain a simple consistent estimator which is a natural counterpart to the CML estimator for standard simultaneous models. An application to a structural model of household labour supply is presented.

Key words: Microeconomics; Censored models; Discrete models
JEL classification: C1; C3; D1

1. Introduction

This paper considers a general class of simultaneous models in which the nonlinearity implicit in the censoring or discrete grouping process prevents an

* Corresponding author.

This research has benefited from the comments of two anonymous referees and from discussions with James Heckman and Arthur van Soest. Research assistance from Alan Duncan and Bruno Henri is gratefully acknowledged. Financial support from the ESRC Centre for the Microeconomic Analysis of Fiscal Policy at the Institute for Fiscal Studies and ESRC grant R000233570 is gratefully acknowledged. We would also like to thank the Department of Employment for use of the Family Expenditure Data. The usual disclaimer applies.

SSDI 0304407693015724

explicit solution for the reduced form. This occurs in the structural shift models of Heckman (1978), the switching equations models described in Gourieroux, Laffont, and Monfort (1980), as well as many applications of joint decision making models on discrete data. The distinction between this class of simultaneous microeconometric models and one in which there is an explicit reduced form depends on whether the structural economic model itself is simultaneous in the *latent* or *observed* dependent variables.[1] This distinction corresponds closely to whether or not the censoring mechanism itself acts as a constraint on individual agents' behaviour.

In the standard class of simultaneous censored or discrete models, individual behaviour is completely described by the latent variable model. The censoring or grouping process simply acts as a constraint on the information available to the econometrician. For example, in a model describing household labour supply and consumption behaviour, hours of work may only be available to the econometrician in grouped form – part-time and full-time – even though individual agents themselves have complete flexibility in their hours' choices. In this case, optimal consumption behaviour is a function of *latent* rather than *observed* labour supply behaviour. As a result the underlying reduced form can be derived explicitly in terms of the latent dependent variables.

In contrast, for the class of models considered in this paper, the observability rule also constrains the agent's choice set. For example, if the two discrete hours packages (part-time, full-time) are all that is available, then household labour supply decisions will reflect this and will depend on the *actual* discrete labour market outcomes rather than the underlying latent hours variables. Similar models arise where there are corner solutions or points of rationing.

Where the model is simultaneous in the underlying latent dependent variables, Blundell and Smith (1989) suggest a conditional maximum likelihood estimator (CMLE) for the structural parameters and considered the relative efficiency properties of this estimator vis-à-vis other consistent estimators. For the class of discrete or censored models of concern here, a further *coherency condition* is required, which imposes restrictions that guarantee the existence of a unique (implicit) reduced form for the *observable* endogenous variables. A discussion of coherency conditions in switching models is provided by Gourieroux, Laffont, and Monfort (1980). For models with discrete dependent variables, a comprehensive discussion is found in Heckman (1978), who refers to this condition as the Principle Assumption.

[1] For this class of models in which there is an explicit reduced form, it is relatively straightforward to construct consistent estimators for the structural parameters based on standard two-stage least squares and instrumental variables approaches. See Amemiya (1978, 1979), Nelson and Olsen (1978), Rivers and Vuong (1988), and Smith and Blundell (1986).

The purpose of this paper is to investigate estimation and inference in this second class of censored or grouped simultaneous equation models. We label these Type IIS simultaneous models. The absence of an explicit (linear) reduced form has a nontrivial impact on the form of the Maximum Likelihood Estimator (MLE). Nevertheless, we derive a simple consistent estimator, which mirrors the CMLE for standard simultaneous models. In addition, this estimator reveals directly how both identification and consistent estimation of structural form parameters depend critically on the satisfaction of appropriate coherency restrictions.

Section 2 describes the models to be considered. These include simultaneous Tobit and Probit models as well as simultaneous extensions of Generalised Selectivity and Double-Hurdle models. In each case, the appropriate coherency conditions and identification are discussed briefly. Our approach to consistent estimation of the structural parameters is described in detail in Section 3. In Section 4 we discuss asymptotic inference, and in Section 5 we present a simple method for discrimination between the two types of models. Section 6 points to the importance of this approach in an application to the estimation of a simultaneous censored model for hours of work decisions. Section 7 concludes the paper.

2. Coherency and identification

The general formulation of our model contains simultaneous Tobit and Probit models as well as simultaneous extensions to the generalised selectivity and the double-hurdle models. The latent dependent variable of interest is denoted by y_{1i}^* with observed value y_{1i}, $i = 1, \ldots, N$. The observation rule linking y_{1i} to y_{1i}^* is given by

$$y_{1i} \equiv g_i(y_{1i}^*, y_{3i}^*), \tag{1}$$

where $g_1(.\,,.)$ is a known function independent of parameters and y_{3i}^* is another latent variable described below which allows (1) to represent selectivity or double-hurdle observability rules on y_{1i}^*. In simple cases such as those of Tobit or Probit, y_{3i}^* does not enter (1) and $y_{1i} \equiv g_1(y_{1i}^*)$, $i = 1, \ldots, N$. The general structural model for y_{1i}^* is defined by

$$y_{1i}^* = \alpha_1 y_{1i} + \beta_1 y_{2i}^* + x_{1i}'\gamma_1 + u_{1i}, \tag{2}$$

where y_{1i} is included to complete the class of models that permit coherency and x_{1i} denotes a k_1-vector of exogenous variables. In (2), y_{2i}^* is a jointly endogenous variable, which is assumed to be continuously observed, that is, $y_{2i} \equiv y_{2i}^*$. For example, in a labour supply context, y_{1i} might represent censored hours of work or discrete participation and y_{2i}^* other household income, jointly determined with labour supply.

The dependent variable y_{2i}^* also depends on the *observed* value y_{1i}:

$$y_{2i}^* = \alpha_2 y_{1i} + \boldsymbol{x}_{2i}'\gamma_2 + u_{2i}, \tag{3}$$

where $\boldsymbol{x}_{2i}$ is a k_2-vector of exogenous variables; both $\boldsymbol{x}_{1i}$ and $\boldsymbol{x}_{2i}$ are assumed to be distributed independently of the error terms u_{1i} and u_{2i}, $i = 1, \ldots, N$. The introduction of y_{1i} rather than its latent value in (3) is the distinguishing characteristic of the type of the Type IIS simultaneous models we consider in this paper. The direct dependence of y_{2i}^* on y_{1i} implies that y_{1i} rather than y_{1i}^* enters the individual decision rule for y_{2i}^*. As a result the observability rule (1) acts as a constraint on individual behaviour as well as a condition for data generation. In the labour supply example, this is akin to including actual hours of work or participation among the determinants of other income.

For completeness, we introduce the following additional latent process y_{3i}^* to allow for more general forms of selectivity:

$$y_{3i}^* = \boldsymbol{x}_i'\pi_3 + v_{3i}, \tag{4}$$

where the k-vector of exogenous variables $\boldsymbol{x}_i$ comprises the nonoverlapping variables of $\boldsymbol{x}_{1i}$ and $\boldsymbol{x}_{2i}$ and is also assumed to be generated independently of v_{3i}, $i = 1, \ldots, N$. The rule linking the observed value y_{3i} to the latent variable y_{3i}^* is denoted by $y_{3i} \equiv g_3(y_{3i}^*)$. This equation is assumed not to depend directly on y_{1i}, and consequently there is no need for a coherency condition in defining the unique reduced form equation (4). In the Generalised Selectivity model y_{3i} will represent the endogenous selection mechanism, while in the Double-Hurdle model it will represent the first hurdle.

This general class of model nests the specifications considered in Heckman (1978) and Gourieroux, Laffont, and Monfort (1980). Finally, we assume the disturbance terms u_{1i}, u_{2i}, and v_{3i} are jointly normally distributed with mean zero and positive definite variance matrix $\Sigma = [\sigma_{ij}]$ and the observations y_{ji}, $j = 1, 2, 3$, and $\boldsymbol{x}_i$, $i = 1, \ldots, N$, constitute a random sample.

Before turning to a detailed analysis of consistent estimation and asymptotic inference in Sections 3 and 4, we review identification and coherency conditions for various common models as special cases of the system described above. To help this discussion, substitute (1) and (3) into (2), which gives

$$y_{1i}^* = (\alpha_1 + \alpha_2\beta_1)g_1(y_{1i}^*, y_{3i}^*) + \boldsymbol{x}_{2i}'\gamma_2\beta_1 + \boldsymbol{x}_{1i}'\gamma_1 + v_{1i}, \tag{5}$$

where $v_{1i} \equiv u_{1i} + \beta_1 u_{2i}$, $i = 1, \ldots, N$. Defining the binary indicator function by $\mathbf{1}(.)$, we have:

Example 2.1 (Tobit). The observability rule (1) becomes

$$y_{1i} \equiv \mathbf{1}(y_{1i}^* > 0)\, y_{1i}^*. \tag{O1}$$

Consider (5) when $y_{1i}^* > 0$:

$$(1 - \alpha_1 - \alpha_2\beta_1)\,y_{1i}^* = \boldsymbol{x}_{1i}'\gamma_1 + \boldsymbol{x}_{2i}'\gamma_2\beta_1 + v_{1i}, \tag{6}$$

from which we see that the coherency condition $(1 - \alpha_1 - \alpha_2\beta_1) > 0$ is sufficient to guarantee a unique reduced form; see Maddala (1983) and Gourieroux, Laffont, and Monfort (1980). Rewrite (6) as

$$y_{1i}^* = \beta_1^*\,\tilde{y}_{2i} + \gamma_1^{*\prime}\boldsymbol{x}_{1i} + u_{1i}^*, \tag{7}$$

when $y_{1i}^* > 0$, where $\tilde{y}_{2i} \equiv y_{2i}^* - \alpha_2 y_{1i}^*$ and $\beta_1^* \equiv \beta_1/(1 - \alpha_1 - \alpha_2\beta_1)$, $\gamma_1^* \equiv \gamma_1/(1 - \alpha_1 - \alpha_2\beta_1)$, $u_{1i}^* \equiv u_{1i}/(1 - \alpha_1 - \beta_1)$. Although β_1^* and γ_1^* are identified, it is easily seen that, given α_2, neither α_1 nor β_1 are. A suitable identifying constraint is thus required; for example, $\alpha_1 = 0$ or $\alpha_1 + \alpha_2\beta_1 = 0$. The former has the advantage of removing the structural shift in (2) while retaining the simultaneity via the inclusion of $\tilde{y}_{2i}$.

Example 2.2 (Probit). The observability rule (1) is

$$y_{1i} \equiv \mathbf{1}(y_{1i}^* > 0). \tag{O2}$$

Again examining (5), it can be seen that the coherency condition is strengthened in this discrete case requiring $\alpha_1 + \alpha_2\beta_1 = 0$ for the existence of a unique reduced form; see Heckman (1978). Moreover, following the discussion of Example 2.1, this condition is sufficient to identify α_1 and β_1, given α_2. It is also sufficient to ensure that the reduced form equation for y_{1i}^* does not depend on the discrete outcome y_{1i} and, as a result, can be solved explicitly.

Example 2.3 (Generalised Selectivity) (Heckman, 1979; Cogan, 1981). The observability rule (1) is

$$y_{1i} \equiv \mathbf{1}(y_{3i}^* > 0)\,y_{1i}^*, \tag{O3}$$

where the dependent variable y_{3i}^* is only observed up to sign, that is $y_{3i} \equiv \mathbf{1}(y_{3i}^* > 0)$. When $y_{3i}^* > 0$, (2) reduces to (6) as in the Tobit model. Thus, both coherency and identification conditions for the Tobit Example 2.1 hold here.

Example 2.4 (Double-Hurdle) (Cragg, 1971; Blundell, Ham, and Meghir, 1987). The observability rule (1) is

$$y_{1i} \equiv \mathbf{1}(y_{1i}^* > 0, y_{3i}^* > 0)\,y_{1i}^*, \tag{O4}$$

with a similar rule to that of Example 2.3 for y_{3i}^*. Again, coherency and identification conditions are as in Example 2.1, because (6) holds when $y_{1i}^* > 0$, $y_{3i}^* > 0$.

3. A consistent two-step estimator

Following the conditional maximum likelihood (CML) approach of Smith and Blundell (1986), we rewrite (2) conditionally on u_{2i} as

$$y_{1i}^* = \alpha_1 y_{1i} + \beta_1 y_{2i}^* + x_{1i}'\gamma_1 + \rho_1 u_{2i} + \varepsilon_{1i}, \tag{8}$$

where $\rho_1 \equiv \sigma_{21}/\sigma_{22}$ and, due to the joint normality of u_{1i} and u_{2i}, $\varepsilon_{1i} \equiv u_{1i} - \rho_1 u_{2i}$ is independent of u_{2i}, $i = 1, \ldots, N$. However, in contrast to the standard simultaneous model in which we can condition directly on the reduced form residual for y_{2i}, it is the constructed variable $\tilde{y}_{2i} \equiv y_{2i}^* - \alpha_2 y_{1i}$ not y_{2i}^* which is independent of ε_{1i} in this problem. Thus, we rewrite (8) as

$$y_{1i}^* = (\alpha_1 + \alpha_2\beta_1) y_{1i} + \beta_1 \tilde{y}_{2i} + x_{1i}'\gamma_1 + \rho_1 u_{2i} + \varepsilon_{1i}, \tag{9}$$

which forms the basis of the suggested estimation procedure for the structural parameters of (2).

To implement (9), it is necessary to obtain suitable estimators of $\tilde{y}_{2i}$ and u_{2i} (or α_2 and γ_2). Consider instrumental variable (IV) estimation of the model for y_{2i}(3). Given sufficient exclusion restrictions on x_i to form x_{2i} $(k > k_2)$, α_2 and γ_2 will be identified by the standard conditions. Although $y_{1i} \equiv g_1(y_{1i}^*, y_{3i}^*)$ is a nonlinear function of the latent variables in (2), under standard conditions, it can be shown that α_2 and γ_2 are consistently estimated by IV using x_i as instruments. Denote these estimators by $\hat{\alpha}_2$ and $\hat{\gamma}_2$ respectively and the corresponding estimated constructed variable and IV residual by $\hat{\tilde{y}}_{2i}$ and $\hat{u}_{2i}$ respectively. Thus, from (9) the relevant estimating equation becomes

$$y_{1i}^* = (\alpha_1 + \alpha_2\beta_1) y_{1i} + \beta_1 \hat{\tilde{y}}_{2i} + x_{1i}'\gamma_1 + \rho_1 \hat{u}_{2i} + \hat{\varepsilon}_{1i}, \tag{10}$$

where

$$\hat{\varepsilon}_{1i} \equiv \varepsilon_{1i} + \beta_1(\tilde{y}_{2i} - \hat{\tilde{y}}_{2i}) + \rho_1(u_{2i} - \hat{u}_{2i}), \qquad i = 1, \ldots, N.$$

Standard ML methods of estimation may now be applied to (10), subject to coherency and identification conditions, [in conjunction with (4) if necessary] to provide consistent estimators of the structural parameters; a formal statement of the consistency result is given in Proposition 3.1 at the end of this section. Intuitively, this result follows since the influence of the second and third terms in the expression for $\hat{\varepsilon}_{1i}$ are asymptotically negligible for the consistency property although these terms will affect the limiting distribution of the resultant estimators; see Section 4.

We now turn to implement the above estimation procedure for the various examples discussed in Section 2:

Example 2.1 (Tobit). Consider (10) for $y_{1i}^* > 0$. After re-arrangement, we obtain

$$(1 - \alpha_1 - \alpha_2\beta_1) y_{1i}^* = \beta_1 \hat{\tilde{y}}_{2i} + x_{1i}'\gamma_1 + \rho_1 \hat{u}_{2i} + \hat{\varepsilon}_{1i},$$

or

$$y_{1i}^* = \beta_1^* \hat{\tilde{y}}_{2i} + x_{1i}'\gamma_1^* + \rho_1^* \hat{u}_{2i} + \hat{\varepsilon}_{1i}^*, \tag{11}$$

where β_1^*, γ_1^* are defined below (7), $\rho_1^* \equiv \rho_1/(1-\alpha_1-\alpha_2\beta_1)$ and $\hat{\varepsilon}_{1i}^* \equiv \hat{\varepsilon}_{1i}/(1-\alpha_1-\alpha_2\beta_1)$. Standard Tobit ML on (11) provides consistent estimators $\hat{\beta}_1^*$, $\hat{\gamma}_1^*$, and $\hat{\rho}_1^*$ for β_1^*, γ_1^*, and ρ_1^*.

Example 2.2 (Probit). Imposing the coherency condition $\alpha_1 + \alpha_2\beta_1 = 0$ on (11) gives

$$y_{1i}^* = \beta_1 \hat{\tilde{y}}_{2i} + x_{1i}'\gamma_1 + \rho_1 \hat{u}_{2i} + \hat{\varepsilon}_{1i}, \tag{12}$$

which may be estimated by standard Probit ML to yield the consistent estimators $\hat{\beta}_1$, $\hat{\gamma}_1$ and $\hat{\rho}_1$ of the structural parameters.

Example 2.3 (Generalised Selectivity). When $y_{3i}^* > 0$, (11) is reproduced. It is necessary in this case also to express Eq. (4) to y_{3i}^* conditionally on u_{2i}, viz.:

$$y_{3i}^* = x_i'\pi_3 + \rho_3 u_{2i} + \varepsilon_{3i}, \tag{13}$$

where $\rho_3 \equiv \sigma_{23}/\sigma_{22}$ and, by joint normality, $\varepsilon_{3i} \equiv v_{3i} - \rho_3 u_{2i}$ is rendered independent of u_{2i} and thus $\tilde{y}_{2i}$, $i = 1, \ldots, N$. Proceeding as for the Tobit case of Example 2.1 above by replacing u_{2i} by the IV residual $\hat{u}_{2i}$ gives

$$y_{3i}^* = x_i'\pi_3 + \rho_3 \hat{u}_{2i} + \hat{\varepsilon}_{3i}, \tag{14}$$

where $\hat{\varepsilon}_{3i} \equiv \varepsilon_{3i} + \rho_2(u_{2i} - \hat{u}_{2i})$, $i = 1, \ldots, N$. Hence, a ML estimation procedure for selectivity models applied to (11) and (14) is appropriate to obtain consistent estimators $\hat{\beta}_1^*$, $\hat{\gamma}_1^*$, and $\hat{\rho}_1^*$ for β_1^*, γ_1^*, and ρ_1^*. Alternatively, the Heckman (1979) procedure applied to (11) would also yield consistent estimators for β_1^*, γ_1^*, and ρ_1^*.

Example 2.4 (Double-Hurdle). The approach discussed in Example 2.3 above applies except that a ML estimation procedure appropriate for the Double-Hurdle model is used.

From the above discussion, cf. (11) and (14), the generic system to be considered is given by

$$y_{1i}^* = \beta_1^* \tilde{y}_{2i} + x_{1i}'\gamma_1^* + \rho_1^* u_{2i} + \varepsilon_{1i}^*, \tag{15}$$

$$y_{3i}^* = x_i'\pi_3 + \rho_3 u_{2i} + \varepsilon_{3i}, \qquad i = 1, \ldots, N, \tag{16}$$

subject to (1); note the dependence of $\tilde{y}_{2i}$ and u_{2i} on $\phi \equiv (\alpha_2', \pi_2')'$. Define $\theta^* \equiv (\beta_1^*, \gamma_1^{*\prime}, \rho_1^*, \pi_3, \rho_3, \Sigma_{**})'$, where $\Sigma_{**}(\equiv \mathrm{var}(\varepsilon_{1i}^*, \varepsilon_{3i})) = S_*(\Sigma_{..} - \sigma_{.}\sigma_{.}'/\sigma_{22})\ S_*$, $S_* \equiv \mathrm{diag}[1/(1-\alpha_1-\alpha_2\beta_1), 1]$, $\sigma_{.} \equiv (\sigma_{12}, \sigma_{32})')'$ and $\Sigma_{..} \equiv (\sigma_{ij}, i, j = 1, 3)$.

Let ϕ_0 and θ_0^* denote the true values of ϕ and θ^*, $\ln \mathscr{L}_N(\theta^*, \phi)$ the log-likelihood for the system (15) and (16) conditional on x_i and u_{2i}, $i = 1, \ldots, N$, and define $\hat{\theta}_N^* \equiv (\hat{\beta}_{1N}^*, \hat{\gamma}_{1N}^*, \hat{\rho}_{1N}^*, \hat{\pi}_{3N}, \hat{\rho}_{3N}, \hat{\Sigma}_{**N})$, the ML estimator for the parameters of (11) and (14); that is, (15) and (16) after substitution of $\hat{\phi}_N \equiv (\hat{\alpha}_{2N}, \hat{\gamma}'_{2N})'$, the IV estimator for $\phi \equiv (\alpha'_2, \gamma'_2)'$.

Proposition 3.1. Let $\hat{\phi}_N$ and $N^{-1} \ln \mathscr{L}_N(\theta^*, \phi)$ satisfy Assumption A.1 of Appendix A. Then, $\hat{\theta}_N^*$ is a weakly consistent estimator for θ_0^*.

Proof. See Theorem A.2 of Appendix A.

Given suitable identifiability constraints linking $\theta_1^* \equiv (\beta_1^*, \gamma_1^{*\prime}, \rho_1^*)'$ and $\theta_1 \equiv (\beta_1, \gamma_1^*, \rho_1)'$, a consistent estimator for θ_1 may be recovered; see Section 4 for an explicit presentation.

The approach described above contrasts directly to the CMLE approach taken by Smith and Blundell (1986) and Blundell and Smith (1989) for standard simultaneous censored or discrete models. In those models, which we label Type IS, an explicit reduced form for y_{2i}^* is estimable at the first step and the Conditional Maximum Likelihood Estimator is obtained by simply conditioning on the residual from the reduced form for y_{2i}. Note that these Type IS models are *not* generally equivalent to setting $\alpha_2 = 0$ in Eq. (3). Although both classes of models are formulated conditionally on a set of exogenous variables $\boldsymbol{x}_i$, in the Type IIS models considered here, it is necessary to provide some exclusion restrictions on $\boldsymbol{x}_i$ to identify the parameters α_2 and γ_2. That is, Type IS and IIS simultaneous models are *nonnested.* A simple method of discriminating between them is presented in Section 5.2 below.

4. Asymptotic inference

The limiting distribution of $\hat{\theta}_N^*$ may be obtained straightforwardly by applying the Amemiya (1978, 1979) approach exploited in Smith and Blundell (1986). Define:

$$\mathscr{I} \equiv -\lim_{N\to\infty} N^{-1} \mathrm{E}[\partial^2 \ln \mathscr{L}_N / \partial\theta^* \partial\theta^{*\prime}],$$

$$\boldsymbol{K} \equiv \mathscr{I}^{-1}\left\{-\lim_{N\to\infty} N^{-1} \mathrm{E}[\partial^2 \ln \mathscr{L}_N / \partial\theta^* \partial\phi']\right\},$$

explicit formulae for which are presented in Appendix B; E[.] denotes expectation taken conditional on $\boldsymbol{x}_i$ and u_{2i}, $i = 1, \ldots, N$. Therefore:

Theorem 4.1. Let $\hat{\phi}_N = (\hat{\alpha}_{2N}, \hat{\gamma}'_{2N})'$ denote the IV estimator for ϕ; thus:

$$N^{1/2}(\hat{\theta}_N^* - \theta_N^*) \xrightarrow{L} N(0, V(\hat{\theta}_N^*)),$$

where

$$V(\hat{\theta}_N^*) = \mathscr{I}^{-1} + \boldsymbol{K}V(\hat{\phi}_N)\boldsymbol{K}',$$

$$V(\hat{\phi}_N)\ (\equiv \text{avar}[N^{1/2}(\hat{\phi}_N - \phi_0)]) = \sigma_{22}(M_{z_*X}M_{XX}^{-1}M_{XZ_*})^{-1},$$

$$M_{Z_*X} \equiv \lim_{T\to\infty} N^{-1}\sum_{i=1}^{N} z_{*i}x_i', \quad etc.$$

and

$$z_{*i} \equiv (y_{1i}, x_{2i}')', \qquad i = 1,\ldots,N.$$

Proof. See Appendix B.

Note that the above limiting distribution result is all that is required to undertake inference for exclusion restrictions on $\theta_1 \equiv (\beta_1, \gamma_1', \rho_1)'$ in Examples 2.1–2.4, given the just identifying assumption $\alpha_1 = 0$ or, trivially, $\alpha_1 + \alpha_2\beta_1 = 0$.

In Examples 2.1, 2.3, and 2.4, having obtained $\hat{\theta}_N^*$, it is still necessary to derive a suitable estimator for θ_1. Given conditions that just identify θ_1 from $\theta_1^* \equiv (\beta_1^*, \gamma_1^{*\prime}, \rho_1^*)'$ such as discussed in those Examples, this matter is relatively straightforward. Consider the following set of constraints linking θ^*, θ, and ϕ:

$$\boldsymbol{q}(\theta^*, \theta, \phi) = \boldsymbol{0},$$

where the number of restrictions equals the number of elements comprising *both* θ^* and θ; furthermore, as above, assume that θ is just identified from θ^* for given ϕ through these constraints. Hence, the required estimator $\hat{\theta}$ for θ is determined uniquely by

$$\boldsymbol{q}(\hat{\theta}_N^*, \hat{\theta}_N, \hat{\phi}_N) = \boldsymbol{0}.$$

As both $\boldsymbol{Q}_\theta\ (\equiv \partial\boldsymbol{q}/\partial\theta')$ and $\boldsymbol{Q}_{\theta^*}\ (\equiv \partial\boldsymbol{q}/\partial\theta^{*\prime})$ are nonsingular, we have, following Szroeter (1983):

$$N^{1/2}(\hat{\theta}_N - \theta_0) \xrightarrow{L} N(\boldsymbol{0}, \boldsymbol{Q}\boldsymbol{V}\boldsymbol{Q}'),$$

where $\boldsymbol{Q} \equiv \boldsymbol{Q}_\theta^{-1}[\boldsymbol{Q}_{\theta^*}, \boldsymbol{Q}_\phi]$, $\boldsymbol{Q}_\phi \equiv \partial\boldsymbol{q}/\partial\phi'$ and

$$\boldsymbol{V} \equiv \begin{pmatrix} \boldsymbol{I}_{\dim(\theta)} & -\boldsymbol{K} \\ \boldsymbol{0} & \boldsymbol{I}_{\dim(\phi)} \end{pmatrix} \begin{pmatrix} \mathscr{I}^{-1} & \boldsymbol{0} \\ \boldsymbol{0} & V(\hat{\phi}_N) \end{pmatrix} \begin{pmatrix} \boldsymbol{I}_{\dim(\theta)} & \boldsymbol{0} \\ -\boldsymbol{K}' & \boldsymbol{I}_{\dim(\phi)} \end{pmatrix},$$

where $\mathscr{I}$ and $\boldsymbol{K}$ are defined above. This latter expression for $\boldsymbol{V}$ ($\equiv$ avar $[N^{1/2}(\hat{\theta}_N^* - \theta_0^*),\ N^{1/2}(\hat{\phi}_N - \phi_0)]$) is obtained from noting that $N^{-1/2}\partial \ln \mathscr{L}_N/\partial\theta^*$ and $N^{1/2}(\hat{\phi}_N - \phi_0)$ are uncorrelated; see Appendix B.

For Examples 2.1, 2.3, and 2.4, the above analysis produces particularly convenient results. Firstly, under the assumption that $\alpha_1 = 0$,

$$\hat{\beta}_{1N} = \hat{\beta}^*_{1N}/(1 + \hat{\alpha}_{2N}\hat{\beta}^*_{1N}),$$

$$\hat{\gamma}_{1N} = \hat{\gamma}^*_{1N}/(1 + \hat{\alpha}_{2N}\hat{\beta}^*_{1N}),$$

$$\hat{\rho}_{1N} = \hat{\rho}^*_{1N}/(1 + \hat{\alpha}_{2N}\hat{\beta}^*_{1N}),$$

and, secondly, defining $\theta_1 \equiv (\beta_1, \gamma_1', \rho_1)'$ as above,

$$N^{1/2}(\hat{\theta}_{1N} - \theta_{10}) \xrightarrow{L} N(0, \boldsymbol{Q}_1 \boldsymbol{V} \boldsymbol{Q}_1'),$$

where $\boldsymbol{Q}_1 \equiv [(1 - \alpha_2\beta_1)\boldsymbol{Q}^{11}_{\theta\theta}, \boldsymbol{0}_{(k_1+2,k+4)}, \beta_1\theta_1, \boldsymbol{0}_{(k_1+2,k_2)}]$ with

$$\boldsymbol{Q}^{11}_{\theta\theta} \equiv -\begin{pmatrix} 1 - \alpha_2\beta_1 & \boldsymbol{0}' & 0 \\ -\alpha_2\gamma_1 & \boldsymbol{I}_{k_1} & \boldsymbol{0} \\ -\alpha_2\rho_1 & \boldsymbol{0}' & 0 \end{pmatrix}.$$

5. Generalisations and tests

5.1. Additional censored regressors

Additional dependent censored regressors may easily be absorbed into the preceding analysis with little change to the foregoing analysis. Consider (2) but respecify it to allow a dependence on the observed y_{3i}. Thus,

$$\begin{aligned} y^*_{1i} &= \alpha_1 y_{1i} + \delta_1 y_{3i} + \beta_1 y^*_{2i} + \boldsymbol{x}'_{1i}\gamma_1 + u_{1i} \\ &= (\alpha_1 + \alpha_2\beta_1)y_{1i} + (\delta_1 + \delta_2\beta_1)y_{3i} + \beta_1\tilde{y}_{2i} \\ &\quad + \boldsymbol{x}'_{1i}\gamma_1 + \rho_1 u_{2i} + \varepsilon_{1i}, \end{aligned} \tag{17}$$

where $\tilde{y}_{2i} \equiv y^*_{2i} - \alpha_2 y_{1i} - \delta_2 y_{3i}$ and (3) is altered in a similar manner, viz.:

$$y^*_{2i} = \alpha_2 y_{1i} + \delta_2 y_{3i} + \boldsymbol{x}'_{2i}\gamma_2 + u_{2i}. \tag{18}$$

For example, if the observability rule determining y_{3i} is that of an indicator function as in Examples 2.3 and 2.4, the additional coherency condition is $\delta_1 + \delta_2\beta_1 = 0$, which simplifies estimation of the conditional model, (14) and (17) after substitution of the IV estimators for the parameter $\phi \equiv (\alpha_2, \delta_2, \gamma_2')$ of (18). The above estimation and inference methodology may be applied with relatively little alteration.

5.2. *Discriminating between Type IS and Type IIS simultaneous models*

Under the null hypothesis that the standard Type IS simultaneous model is correct, the reduced form for y_{2i} may be expressed as $y_{2i} = x_i'\pi_2 + v_{2i}$, $i = 1, \dots, N$. Thus, one can test the Type IS against the Type IIS specification by estimating

$$y_{2i} = x_i'\pi_2 + \hat{y}_{1i}\delta_1 + \xi_{1i}, \tag{19}$$

by least squares and testing $\delta_1 = 0$, where $\hat{y}_{1i}$ is the estimated prediction for y_{1i} from the Type IIS specification (5).[2] Alternatively, under the Type IIS null given in (3), the Type IIS specification may be tested against the Type IS model by estimating

$$y_{2i} = y_{1i}\alpha_2 + x_{2i}'\gamma_2 + \hat{y}_{2i}\delta_2 + \xi_{2i}, \tag{20}$$

by instrumental variables using x_i as instruments and testing $\delta_2 = 0$, where $\hat{y}_{2i}$ is the estimated prediction of y_{2i} under the Type IS explicit linear reduced form for y_{2i} given above. See Davidson and MacKinnon (1981) and Godfrey (1983).

6. An empirical application

In the application, we consider the case of a joint decision making model for married women's hours of work and other household income. This serves as a useful application since other household income is continuously observed where as female hours of work may sensibly be subject to any of the four observability rules (O1)–(O4) described in Section 2. For example, the standard classical 'corner solution' model of hours of work and participation is described by the Tobit observability rule (O1) (see Killingsworth, 1993, Ch. 3). This is often extended to the Generalised Selectivity model (O3) if either wages are not observed for nonparticipants (Heckman, 1979) or where fixed costs of work break the relation between hours of work and participation (Cogan, 1981). If hours of work are not observed, then a simple binary Probit (O2) model may be adopted to describe participation. Finally, if some nonworkers would like to work but cannot obtain employment, we have the Double-Hurdle model (O4) for labour supply as described in Blundell, Ham, and Meghir (1987), where the first hurdle relates to job availability.

[2] For example, the Tobit specification (7) gives $y_{1i}^* = x_{1i}'\gamma_1^* + x_{2i}'\gamma_2\beta_1^* + v_{1i}^*$, where $v_{1i}^* \equiv u_{1i}^* + \beta_1^* u_{2i}$, for $y_{1i}^* > 0$. Thus $\mathrm{E}[y_{1i}|x_i] = \Phi[(x_{1i}'\gamma_1^* + x_{2i}'\gamma_2\beta_1^*)/\omega^*)]\ (x_{1i}'\gamma_1^* + x_{2i}'\gamma_2\beta_1^*) + \omega^*\phi[(x_{1i}'\gamma_1^* + x_{2i}'\gamma_2\beta_1^*)/\omega^*]$, where $\omega^{*2} \equiv \mathrm{var}[v_{1i}^*]$, $\phi(.)$ and $\Phi(.)$ denote the standard normal density and distribution functions respectively.

In each of these models the determinants of the other income variable will contain the labour supply decisions of other household members as well as any household savings decisions. If these decisions are jointly determined with female hours of work, then they will depend on the *actual* hours of work not the underlying latent desired hours variable. Thus, the model will most likely fit into the Type II Simultaneous (Type IIS) framework described earlier.

Therefore, notationally, y_{1i} in (2) relates to the hours of work while y_{2i} refers to the other income variable.[3] The data, details of which are provided in the Data Appendix, are drawn from the UK Family Expenditure Survey for 1981. The hours of work variable refers to the normal weekly hours of work for a sample of (2539) married women of working age for whom the sample participation rate is 56%. The exogenous factors x_{1i} that determine hours of work relate to the age of the woman (a_f), the presence of a young child (D_1), and her years of education (e_f). The exogenous variables in the other income equation in x_{2i} include the age, demographic, and education variables. In addition housing tenure dummies (T_i), husband's skill characteristics (MO_i), and the local unemployment rate (UN) are added to identify the model. Brief summary statistics for all these variables are presented in the Data Appendix.

Table 1 presents the instrumental variable estimates of the other income equation corresponding to (3) in Section 2. Notice that α_2 is negative and significant. Following the procedure outlined in Section 3, we estimate the structural parameters of the censored hours equation conditional on $\hat{\alpha}_2$ and $\hat{u}_{2i}$, using the standard Tobit estimator applied to (10). These results are presented in Table 2, where we have recovered the underlying structural parameters β_1, γ_1, and ρ_1 from the β_1^*, γ_1^*, and ρ_1^* estimates, using the identification condition $\alpha_1 = 0$. The coherency condition then simply reduces to $\alpha_2\beta_1 < 1$ which, given our estimate for β_1 in Table 2, is seen to be satisfied.

Although the argument for including *actual* hours y_{1i} rather than *desired* hours y_{1i}^* as an explanatory factor for other income y_{2i} is convincing when y_{1i} and y_{2i} are jointly determined, it is quite possible that the Type IS model may be more appropriate in which case there is an explicit reduced form for y_{2i}. In particular, the model has a recursive structure and the conditional model simply involves the inclusion of the reduced form residual, $\hat{v}_{2i} = y_{2i} - x_i'\hat{\pi}_2$, in the structural equation for y_{1i} (Smith and Blundell, 1986). As described in Section 5.2 above, the two classes of models are strictly nonnested and may be compared using the statistics from (19) and (20); these results are given after presenting a comparison of the estimation results for each model.

To provide a reference point in the comparison, the first column in Table 3 contains the standard Tobit estimates for the model of Table 2. A comparison

[3] This we define in the life-cycle consistent manner as developed in Blundell and Walker (1986).

Table 1
The other income model

	Parameter estimate	Standard error	Variable mean
y_1	− 0.4659	0.1803	15.161
a_f	4.9003	1.3108	− 0.401
n_1	3.8115	3.7197	0.391
n_2	6.9561	1.8189	0.467
T_1	− 7.2911	1.9492	0.285
T_2	− 5.2239	3.8164	0.046
MO_1	− 12.2799	3.0812	0.080
MO_2	− 10.8431	1.9086	0.392
MO_3	− 14.5201	2.4212	0.175
UN	− 0.5017	0.2722	13.482
Const.	77.7339	7.2661	1.00

Exact definitions of variables in the Data Appendix. All calculations were performed using GAUSS-386.

Table 2
Hours of work and the Type IIS specification

	Parameter estimate	Standard error	Variable mean
$\tilde{y}_2$	− 0.0929	0.0222	50.525
a_f	− 2.0437	0.3257	− 0.401
a_f^2	0.5391	0.2506	1.282
e_f	0.6762	0.1182	2.928
e_f^2	− 0.0164	0.0114	14.825
D_1	− 24.0999	1.5226	0.295
$\hat{u}_2$	0.2104	0.0509	0.00
Const.	25.1666	2.1531	1.00

Exact definitions of variables in the Data Appendix. All calculations were performed using GAUSS-386.

with the estimates in Table 2 indicates the degree of bias involved in assuming y_{2i} to be exogenous in the determination of y_{1i}. The second column in Table 3 presents the estimation results from the conditional model assuming y_{1i}^* enters the determination of y_{2i}, $\hat{v}_{2i}$ is then the reduced form error term where all x_i are used as instruments. Table 3 suggests that incorrectly assuming the standard simultaneous model could lead to large overadjustment for simultaneous equations bias. However, that conclusion rests on the assumption that the Type IIS structure is the correct specification. To assess this, the t-value for $\delta_1 = 0$ in (19) is 3.379, whereas that for $\delta_2 = 0$ in (20) has the value 1.712. This provides some

Table 3
The Tobit and Type IS specifications

	Tobit	Type IS
y_2	−0.1211 (0.0109)	−0.0477 (0.0291)
a_f	−3.8571 (0.4407)	−5.4313 (0.6379)
a_f^2	0.9234 (0.4166)	0.8332 (0.4252)
e_f	0.7082 (0.1769)	0.6545 (0.1834)
e_f^2	−0.0089 (0.0171)	−0.0258 (0.0179)
D_1	−24.7218 (1.2545)	−28.7210 (2.2119)
$\hat{v}_2$	...	−0.1251 (0.0491)
Const.	29.0583 (1.1996)	25.2148 (1.8434)

Standard errors in parentheses.

evidence in favour of the Type IIS model for these data. A comparison of the Type IIS and Type IS results suggests that the standard simultaneous Tobit model (Type IS) overadjusts for simultaneity bias.

7. Conclusions

A simple consistent estimator is proposed for a class of simultaneous microeconometric models in which censoring or grouping of the dependent variables implies the lack of an explicit reduced form, thus requiring a coherency condition. Our estimator corresponds to the conditional maximum likelihood estimator proposed in Blundell and Smith (1989) for an alternative class of simultaneous models in which no coherency condition is needed and an explicit (linear) reduced form may be derived. We show that this estimation procedure can be applied across a wide variety of popular models and that it provides a useful basis for inference in such models. Finally, we apply this methodology to a model of the joint determination of hours of work and other household income for a sample of married couples in the UK.

Appendix A

Consider the following assumption adapted from Amemiya (1985, Theorem 4.1.1, pp. 106–107):

Assumption A.1. (i) $\hat{\phi}_N \xrightarrow{P} \phi_0$. (ii) (a) $\mathcal{Q}_N(\theta^*, \phi) \xrightarrow{P} \mathcal{Q}(\theta^*, \phi)$ *uniformly in* $(\theta^*, \phi) \in \Theta^* \times \Phi$, *assumed compact, as* $N \to \infty$; (b) $\mathcal{Q}(\theta^*, \phi)$ *attains a unique global maximum at* (θ_0^*, ϕ_0). (iii) $\mathcal{Q}_N(\theta^*, \phi)$ *is continuous in* (θ^*, ϕ) *and measurable.*

Remark. Alternative conditions to those given in Assumption A.1(ii), (iii) may be obtained by using the results of Andrews (1987) and Newey (1991).

Define $\hat{\theta}_N^*$ as a value that satisfies

$$\mathcal{Q}_N(\hat{\theta}_N^*, \hat{\phi}_N) = \max_{\theta^* \in \Theta^*} \mathcal{Q}_N(\theta^*, \hat{\phi}_N). \tag{A.1}$$

Theorem A.2. Let $\hat{\phi}_N$ *and* $\mathcal{Q}_N(\theta^*, \phi)$ *satisfy Assumption A.1. Then:*

$$\hat{\theta}_N^* \xrightarrow{P} \theta_0^*.$$

Proof. Define $\mathcal{N}$ as an open neighbourhood of θ_0^*; thus $\mathcal{N}^c \cap \Theta^*$ is compact and, hence, $\max_{\theta^* \in \mathcal{N}^c \cap \Theta} \mathcal{Q}(\theta^*, \phi_0)$ exists. Let:

$$\varepsilon \equiv \mathcal{Q}(\theta_0^*, \phi_0) - \max_{\theta^* \in \mathcal{N}^c \cap \Theta^*} \mathcal{Q}(\theta^*, \phi_0).$$

Define $\mathcal{A}_N \equiv \{\omega: |\hat{\phi}_N - \phi_0| < \delta\}$, where, using A.1(iii), $\delta > 0$ is chosen such that $|\mathcal{Q}_N(\theta^*, \hat{\phi}_N) - \mathcal{Q}_N(\theta^*, \phi_0)| < \varepsilon/4$, $\varepsilon > 0$. Furthermore, define $\mathcal{B}_N \equiv \{\omega: |\mathcal{Q}_N(\theta^*, \hat{\phi}_N) - \mathcal{Q}_N(\theta^*, \phi_0)| < \varepsilon/4\}$. Note $\mathcal{A}_N$ implies $\mathcal{B}_N$; hence, $P[\mathcal{B}_N] \to 1$ as $N \to \infty$ from A.1(i). Now, $\mathcal{B}_N$ implies

$$\mathcal{Q}_N(\hat{\theta}_N^*, \phi_0) > \mathcal{Q}_N(\hat{\theta}_N^*, \hat{\phi}_N) - \varepsilon/4, \tag{A.2}$$

$$\mathcal{Q}_N(\theta_0^*, \hat{\phi}_N) > \mathcal{Q}_N(\theta_0^*, \phi_0) - \varepsilon/4, \tag{A.3}$$

and from (A.1),

$$\mathcal{Q}_N(\hat{\theta}_N^*, \hat{\phi}_N) \geqslant \mathcal{Q}_N(\theta_0^*, \hat{\phi}_N). \tag{A.4}$$

Define $\mathcal{C}_N \equiv \{\omega: |\mathcal{Q}_N(\theta^*, \phi) - \mathcal{Q}(\theta^*, \phi)| < \varepsilon/4\}$ and, thus, $P[\mathcal{C}_N] \to 1$ as $N \to \infty$ by Assumption A.1(ii) (a). Now, $\mathcal{C}_N$ implies

$$\mathcal{Q}(\hat{\theta}_N^*, \phi_0) > \mathcal{Q}_N(\hat{\theta}_N^*, \phi_0) - \varepsilon/4, \tag{A.5}$$

$$\mathcal{Q}_N(\theta_0^*, \phi_0) < \mathcal{Q}(\theta_0^*, \phi_0) - \varepsilon/4. \tag{A.6}$$

(A.2) and (A.4) yield

$$\mathcal{Q}_N(\hat{\theta}_N^*, \phi_0) > \mathcal{Q}_N(\theta_N^*, \hat{\phi}_N) - \varepsilon/4. \tag{A.7}$$

(A.3) and (A.7) give

$$\mathcal{Q}_N(\hat{\theta}_N^*, \phi_0) > \mathcal{Q}_N(\theta_N^*, \phi_0) - \varepsilon/2. \tag{A.8}$$

(A.6) and (A.8) yield

$$\mathcal{Q}_N(\hat{\theta}_N^*, \phi_0) > \mathcal{Q}(\theta_0^*, \phi_0) - 3\varepsilon/4. \tag{A.9}$$

Therefore, adding (A.5) and (A.9) gives

$$\mathcal{Q}(\hat{\theta}_N^*, \phi_0) > \mathcal{Q}(\theta_0^*, \phi_0) - \varepsilon. \tag{A.10}$$

Thus, from (A.10), $\mathcal{B}_N \cap \mathcal{C}_N$ implies $\hat{\theta}_N^* \in \mathcal{N}$ or $P[\mathcal{B}_N \cap \mathcal{C}_N] \leqslant P[\hat{\theta}_N^* \in \mathcal{N}]$. As $P[\mathcal{B}_N \cap \mathcal{C}_N] \geqslant 1 - \{P[\mathcal{B}_N^c] + P[\mathcal{C}_N^c]\}$, $P[\mathcal{B}_N \cap \mathcal{C}_N] \to 1$ as $N \to \infty$. Hence, $P[\hat{\theta}_N^* \in \mathcal{N}] \to 1$ as $N \to \infty$.

Appendix B

Consider system (15) and (16) subject to (1) after substitution of $\hat{\phi}_N$. Thus, a first-order Taylor series expansion about (θ_0^*, ϕ_0) yields

$$N^{1/2}(\hat{\theta}_N^* - \theta_0^*) = \mathcal{I}^{-1} N^{-1/2}(\partial \ln \mathcal{L}_N/\partial\theta^*) - \boldsymbol{K} N^{1/2}(\hat{\phi}_N - \phi_0) + o_P(1); \tag{B.1}$$

cf. Amemiya (1978, 1979) and Smith and Blundell (1986).

Following Smith (1987, App. B, pp. 120–121), the score vector for system (15) and (16) subject to (1), assuming that $\tilde{y}_{2i}$ and u_{2i} are observed, $i = 1, \ldots, N$, is given by

$$\partial \ln \mathcal{L}_N/\partial\theta^* = \sum_{i=1}^{N} w_i \zeta_i,$$

under the assumption that $\varepsilon_i (\equiv (\varepsilon_{1i}^*, \varepsilon_{3i})) \sim NI(0, \Sigma_{**})$, $i = 1, \ldots, N$, $\zeta \equiv [\varepsilon', \overline{(\varepsilon \otimes \varepsilon)'}]'$,

$$\boldsymbol{w} \equiv \boldsymbol{S}\begin{pmatrix} \Sigma_{**}^{-1} \otimes z & \boldsymbol{0} \\ \boldsymbol{0} & \frac{1}{2}(\Sigma_{**}^{-1} \otimes \Sigma_{**}^{-1}) \end{pmatrix},$$

$z \equiv (\tilde{y}_2, x', u_2)'$ and $S \equiv \operatorname{diag}(S_1, S_3, D)$; S_1 and S_3 select out the appropriate elements of z included in (15) and (16) respectively, whereas D obeys $D' v(\Sigma) = \operatorname{vec}(\Sigma)$ with $v(.)$ selecting out the distinct elements of a symmetric matrix (Magnus and Neudecker, 1980). The term ζ is defined by the generalised error products discussed in Gourieroux et al. (1987) and Smith (1987); thus:

$$\bar{\varepsilon} \equiv \mathrm{E}[\varepsilon \mid y], \quad \overline{\varepsilon \otimes \varepsilon} \equiv \mathrm{E}[\varepsilon \otimes \varepsilon \mid y] - \operatorname{vec}(\Sigma_{**}),$$

where $\mathrm{E}[.\mid y]$ denotes conditional expectation given the observability rule linking $y^* = (y_1^*, y_3^*)'$ to $y = (y_1, y_3)'$ and z. By the familiar information matrix equality and the assumption of the independence of observations $i = 1, \ldots, N$, we have that conditional on z:

$$-N^{-1}\mathrm{E}[\partial^2 \ln \mathscr{L}_N / \partial\theta^* \partial\theta^{*\prime}] = N^{-1} \sum_{i=1}^{N} w_i \operatorname{var}[\zeta_i]\, w_i'.$$

The score vector of $\ln \mathscr{L}_N$ with respect to $\phi = (\alpha_2, \gamma_2')'$ is

$$\partial \ln \mathscr{L}_N / \partial\phi = -\sum_{i=1}^{N} w_{2i} \bar{\varepsilon}_i,$$

where $w_2 \equiv S_2 (\Sigma_{**}^{-1} \otimes z_2)$, $z_2 \equiv (y_1, x_i')'$, and

$$S_2 \equiv \begin{pmatrix} (\beta_1 + \rho_1) & \mathbf{0}' & \rho_3 & \mathbf{0}' \\ 0 & \rho_1 I_{k_2} & \mathbf{0} & \rho_3 I_{k_2} \end{pmatrix}.$$

Thus, again by independence and the information matrix equality:

$$-N^{-1}\mathrm{E}[\partial^2 \ln \mathscr{L}_N / \partial\theta^* \partial\phi'] = -N^{-1} \sum_{i=1}^{N} w_i \operatorname{cov}[\zeta_i, \bar{\varepsilon}_i]\, w_{2i}';$$

cf. Smith and Blundell (1986).

Now, $\operatorname{avar}[N^{-1/2}\partial \ln l_N / \partial\theta^*, N^{1/2}(\hat{\phi}_N - \phi_0)] = \mathbf{0}$ as

$$N^{1/2}(\hat{\phi}_N - \phi_0) = \sigma_{22}^{-1} V(\hat{\phi}_N) M_{Z_* X} M_{XX}^{-1} N^{-1/2} \sum_{i=1}^{N} z_{*i} u_{2i} + o_P(1),$$

and, conditional on x_i and u_{2i}, $i = 1, \ldots, N$, $\mathrm{E}[\partial \ln \mathscr{L}_N / \partial\theta^*)] = \mathbf{0}$; cf. Smith and Blundell (1986, App.). Therefore, from (B.1):

$$V(\hat{\theta}_N^*) = \mathscr{I}^{-1} + K V(\hat{\phi}_N) K'.$$

Data appendix

The data are a sample of 2539 married women from the 1981 Family Expenditure Survey for the UK. All women are of working age and are not self-employed.

		Mean	Standard deviation
Female hours:	y_1	15.1611	15.7452
Other income:	y_2	50.5254	41.7982
(Age − 40)/10:	a_f	− 0.4008	1.0593
(Age − 40)²/100:	a_f^2	1.2823	1.1580
(Education − 8):	e_f	2.9283	2.5005
(Education − 8)²:	e_f^2	14.8251	25.3807
Youngest kid age [–, 5]:	D_1	0.2954	0.4563
Youngest kid age [5, 10]:	D_2	0.2209	0.4149
Youngest kid age [11, –]:	D_3	0.1394	0.3463
Number of kids [–, 5]:	N_1	0.3911	0.6680
Number of kids [5, 10]:	N_2	0.4667	0.7400
Number of kids [11, –]:	N_3	0.4124	0.7499
Owner occupier:	T_1	0.2847	0.4514
Local authority:	T_2	0.0456	0.2088
Husband: skilled:	MO_1	0.0803	0.2718
Husband: semiskilled:	MO_2	0.3919	0.4882
Husband: unskilled:	MO_3	0.1748	0.3798
Local unemployment:	UN	13.4821	2.9086

y_1 = normal weekly hours of work for married women, y_2 = normal weekly earnings minus total consumption expenditures.

References

Amemiya, T., 1978, The estimation of a simultaneous equation generalised probit model, Econometrica 46, 1193–1205.

Amemiya, T., 1979, The estimation of a simultaneous equation tobit model, International Economic Review 20, 169–181.

Amemiya, T., 1985, Advanced econometrics (Harvard University Press, Cambridge, MA).

Andrews, D.W.K., 1987, Consistency in nonlinear econometric models: A generic uniform law of large numbers, Econometrica 55, 1465–1471.

Blundell, R.W. and I. Walker, 1986, A life-cycle consistent model of family labour supply using cross section data, Review of Economic Studies LIII, 539–558.

Blundell, R.W., J. Ham, and C. Meghir, 1987, Unemployment and female labour supply, Economic Journal 97, 44–64.

Blundell, R.W. and R.J. Smith, 1989, Estimation in a class of limited dependent variable models, Review of Economic Studies 56, 37–58.

Cogan, J.F., 1981, Fixed costs and labour supply, Econometrica 49, 945–964.

Cragg, J.G., 1971, Some statistical models for limited dependent variables with applications to the demand for durable goods, Econometrica 39, 829–844.

Davidson, R. and J.G. MacKinnon, 1981, Several tests for model specification in the presence of alternative hypotheses, Econometrica 49, 781–793.

Godfrey, L.G., 1983, Testing non-nested models after estimation by instrumental variables or least squares, Econometrica 51, 355–366.

Gourieroux, C., J.J. Laffont, and A. Monfort, 1980, Coherency conditions in simultaneous linear equation models with endogenous switching regimes, Econometrica 48, 675–695.

Gourieroux, C., A. Monfort, E. Renault, and A. Trognon, 1987, Generalised residuals, Journal of Econometrics 34, 5–32.

Heckman, J.J., 1978, Dummy endogenous variables in a simultaneous equation system, Econometrica 46, 931–959.

Heckman, J.J., 1979, Sample selection bias as specification error, Econometrica 47, 153–162.

Killingsworth, M., 1983, Labour supply, Cambridge surveys of economic literature (Cambridge University Press, Cambridge).

Maddala, G.S., 1983, Limited-dependent and qualitative variables in econometrics (Cambridge University Press, Cambridge).

Magnus, J.R. and H. Neudecker, 1980, The elimination matrix: Some theorems and applications, SIAM Journal of Algebraic and Discrete Methods 1, 422–449.

Nelson, F. and L. Olsen, 1978, Specification and estimation of a simultaneous equation model with limited dependent variables, International Economic Review 19, 695–705.

Newey, W.K., 1991, Uniform convergence in probability and stochastic equicontinuity, Econometrica 59, 1161–1167.

Rivers, D. and Q.H. Vuong, 1988, Limited information estimators and exogeneity tests for simultaneous probit models, Journal of Econometrics 39, 347–366.

Smith, R.J., 1987, Testing the normality assumption in multivariate simultaneous limited dependent variable models, Journal of Econometrics 34, 105–123.

Smith, R.J. and R.W. Blundell, 1986, An exogeneity test for a simultaneous equation Tobit model with an application to labor supply, Econometrica 54, 679–685.

Szroeter, J., 1983, Generalised Wald methods for testing nonlinear implicit and overidentifying restrictions, Econometrica 51, 335–353.